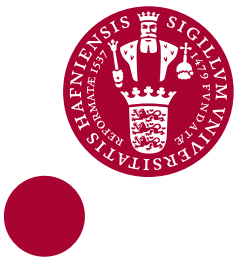

Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression

Pinto, Rita; Hansen, Lars; Hintze, John Birger Hjalmar; Almeida, Raquel; Larsen, Sylvester; Coskun, Mehmet; Davidsen, Johanne; Mitchelmore, Cathy; David, Leonor; Troelsen, Jesper Thorvald; Bennett, Eric Paul

Published in:
Nucleic Acids Research

DOI:
10.1093/nar/gkx371

Publication date:
2017

Document version
Publisher's PDF, also known as Version of record

Published online 4 May 2017

Nucleic Acids Research, 2017, Vol. 45, No. 13 ***e123***
doi: 10.1093/nar/gkx371

Precise integration of inducible transcriptional elements (PrIITE) enables absolute control of gene expression

Rita Pinto[1,2,3,4], Lars Hansen[4], John Hintze[4], Raquel Almeida[1,2,3,5], Sylvester Larsen[6,7], Mehmet Coskun[8,9], Johanne Davidsen[6], Cathy Mitchelmore[6], Leonor David[1,2,3], Jesper Thorvald Troelsen[6] and Eric Paul Bennett[4,*]

[1]i3S, Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal, [2]Ipatimup, Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal, [3]Faculty of Medicine of the University of Porto, Porto, Portugal, [4]Copenhagen Center for Glycomics, Departments of Cellular and Molecular Medicine and Odontology, Faculty of Health Sciences, University of Copenhagen, Copenhagen, Denmark, [5]Department of Biology, Faculty of Sciences of the University of Porto, Porto, Portugal, [6]Department of Science and Environment, Roskilde University, Roskilde, Denmark, [7]Department of Clinical Immunology, Naestved Hospital, Naestved, Denmark, [8]Department of Gastroenterology, Medical Section, Herlev Hospital, University of Copenhagen, Copenhagen, Denmark and [9]The Bioinformatics Centre, Department of Biology & Biotech Research and Innovation Centre (BRIC), University of Copenhagen, Copenhagen, Denmark

Received October 03, 2016; Revised March 30, 2017; Editorial Decision April 10, 2017; Accepted April 27, 2017

ABSTRACT

Tetracycline-based inducible systems provide powerful methods for functional studies where gene expression can be controlled. However, the lack of tight control of the inducible system, leading to leakiness and adverse effects caused by undesirable tetracycline dosage requirements, has proven to be a limitation. Here, we report that the combined use of genome editing tools and last generation Tet-On systems can resolve these issues. Our principle is based on precise integration of inducible transcriptional elements (coined PrIITE) targeted to: (i) exons of an endogenous gene of interest (GOI) and (ii) a safe harbor locus. Using PrIITE cells harboring a GFP reporter or CDX2 transcription factor, we demonstrate discrete inducibility of gene expression with complete abrogation of leakiness. CDX2 PrIITE cells generated by this approach uncovered novel CDX2 downstream effector genes. Our results provide a strategy for characterization of dose-dependent effector functions of essential genes that require absence of endogenous gene expression.

INTRODUCTION

Historically, analysis of the molecular genetic mechanisms underlying cell fate and animal phenotypes has been studied by abrogating gene function in cellular and animal model systems. Initially this has been accomplished by random mutagenesis (1–3), homologous recombination (4) and recently by the use of precise genome editing technologies that allow for target inactivation of any GOI in cells, tissues and animal models (5–7). However, the fact that the genetic lesions induced are static render these approaches inadequate in situations where swift reversal of gene function is desired or in cases where the GOI plays an essential function for cellular survival. Thus, alternative approaches have to be employed in these situations. One commonly used alternative is based on gene 'knock down' by RNAi/shRNA (8). Although the successful application of these technologies in cell lines is well documented, knock down strategies are hampered by lack of quantitative and absolute inactivation of gene function, which makes this approach problematic in situations where downstream gene functional studies require complete gene inactivation. In these situations, the use of inducible gene expression systems has shown to be a powerful methodology that allows for: (i) control of gene expression levels of potentially toxic gene products that could have adverse side effects on cell growth and survival when expressed constitutively, (ii) temporal and spatially controlled activation of genes and proteins and (iii) analysis of cellular gene dose/response effects. Various inducible gene ex-

*To whom correspondence should be addressed. Tel: +45 35326630; Email: epb@sund.ku.dk
Present address: Rita Pinto, Department of Biosciences, Faculty of Mathematics and Natural Sciences, University of Oslo, Blindern, 0316 Oslo, Norway.

pression systems have been described in the literature based on cre-lox P system (9), myxovirus resistance 1 promoter (10), estrogen receptor (11), optogenetics (12), ecdysone-induciblity (13) or tetracyclin (Tet)-Off/On systems (14–17). The latter systems are probably the most commonly used inducible systems for which a plurality of reagents have been developed, published and are commercially available (18). However, one disadvantage of Tet-inducible systems and the majority of the aforementioned inducible systems, is their well described 'leakiness' (19–23). Furthermore, inducible expression in both transient and stable expressing cells have indicated that cells respond differently to induction, which has been attributed to heterogeneity in chromosomal integration of the inducible gene elements in individual cells leading to non-homogenous induction responses (21). Importantly, well known side effects of tetracycline (and its derivatives) on cell fitness, in particular after long term treatment, caution for its use in biomedical research when used at traditional concentrations >100 ng/ml (24–26).

Thus, there is a yet unmet need in the field to improve the tightness of the available inducible systems. We reasoned that the observed 'leakiness' with the most commonly used 'Tet-On' system (17,27) is due to the uncontrolled randomness of integration of the genetic elements encoding both the ectopically expressed transactivator and the inducible GOI, leading to muddled inducible gene expression with pleiotropic downstream effects dependent on the activation conditions used. Therefore, we hypothesize that by integrating a defined number of transactivator and inducible transcriptional elements at defined cellular genomic loci we can (i) circumvent the 'leakiness' issue and (ii) lower the Dox concentration needed for induction below the levels causing cellular stress. Thus, by taking advantage of precise genome editing and last generation Tet-On platforms (18,27), the objectives of this study were to establish a flexible 'non-leaky', minimal Dox concentration requiring isogenic knockout-rescue system. We built our cellular model system on the colorectal cell line LS174T and by mono- or bi-allelic targeting of constitutively expressing transactivator (Tet3G) elements (TET3G) to one locus and inducible GFP-reporter elements to another safe harbor locus we demonstrate minimal Dox requirement, no leakiness and reversibility of the system. We next demonstrate the utility of the isogenic PrIITE system by mono- or bi-allelic integration of inducible CDX2 transcription factor elements into PrIITE cells and re-confirm the reversibility and lack of leakiness of the system. We also for the first time demonstrate, that Tet-On leakiness is related to the cellular copy number of integrated transactivator elements and not the number of integrated inducible elements of the inducible system.

Finally we confirm the complete absence in leakiness of the system by RNA-seq and in combination with ChIP-seq identify several novel genes directly transcriptionally controlled by CDX2 (Figure 1). The selection of CDX2 as our target gene was based on its relevance as a key regulator of intestinal differentiation, with many downstream targets that can be assessed as read-outs (28).

MATERIALS AND METHODS

ZFN gene targeting plasmids and plasmid donor construction

CDX2 and AAVS1 CompoZr ZFN plasmids for human *CDX2* and *AAVS1* were obtained from Sigma targeting exon 1 of *CDX2* (AACTTCGTCAGCCCCccgcagTACCC GGACTACGGCGGTT) and intron one of the PPP1R12C gene at the *AAVS1* hotspot AAV integration site (ACCCCA CAGTGGggccacTAGGGACAGGAT), respectively. ZFN binding sites are shown in upper case and linker cut site in lower case lettering. ZFNs were tagged with 2A peptide fused GFP or Crimson as described recently (29). Fluorescent protein tagging allows for FACS enrichment and improves efficiency in obtaining correctly targeted clones.

Donor construct was designed based on the previously described approach by Maresca *et al.* (30) with the only modification that inverted ZFN binding sites were positioned flanking the entire donor insert. A synthetic ObLiGaRe CDX2 donor vector frame work possessing inverted CDX2 ZFN binding sites (AACCGCCG TAGTCCGGGTAccgcagGGGGCTGACGAAGTT) flanking a XhoI/EcoRV linker was generated (EPB64, Genewiz, USA, Addgene ID#90017). A CMV-Tet3G-SV40UTR TET3G transactivator encoding fragment (pTet3G, Clontech/Takara, USA) was XhoI/HindIII excised (HindIII overhang bluntended) and inserted into XhoI/XbaI (overhangs bluntended) site of EPB64-donor vector generating *pCDX2-pCMV-TET3G-ObLiGaRe* donor vector.

Various inducible gene expression ObLiGaRe donor constructs targeted to the *AAVS1* safe harbor locus were generated based on available *AAVS1* safe harbor CompZr ZFN binding element information (Sigma-Aldrich, USA). A *AAVS1* ObLiGaRe donor vector framework was generated (EPB58, Genewiz, USA, Addgene ID#90016) and designated pObLiGaRe-*AAVS1*, possessing the respective inverted ZFN binding elements (5′-atcctgtcccta ggccacccactgtgggt-3′) flanking an EcoRV multiple cloning site linker (see Supplementary Figure S11 for vector design). A full-length codon optimized CDX2 (CDX2opt) construct was generated (EPB40, GeneArt/ThermoFisher, USA and SalI/BamHI insert cloned into SalI/BamHI sites of pTRE3G (Clontech/Takara, USA) generating pTRE3G-CDX2opt.

A BamHI fragment encoding a Golgi targeted EYFP fusion protein, T2EYFP (31) was blunt end cloned into the EcoRV site of pTRE3G generating *p*TRE3G-T2EYFP. T2EYFP encodes the N-terminal Golgi targeting and retention sequence fused to EYFP. The XbaI/XhoI Tet responsive expression fragments from both *p*TRE3G-CDX2opt and *p*TRE3G-T2EYFP were excised and blunt end cloned into the EcoRV site of pObLiGaRe-AAVS1 (EPB58) generating *p*AAVS1-TRE3G-CDX2opt and *p*AAVS1-TRE3G-T2EYFP. For *p*AAVS1-TRE3G-T2EYFP, the donor insert was excised with PacI/PmeI and inserted into PacI/PmeI site of a modified AAVS1 targeting vector (EPB71, Genewiz, USA, Addgene ID#90018) where PacI/PmeI cloning sites are flanked by insulator sequences and the inverted AAVS1 ZFN binding sites. See Supplementary Fig-

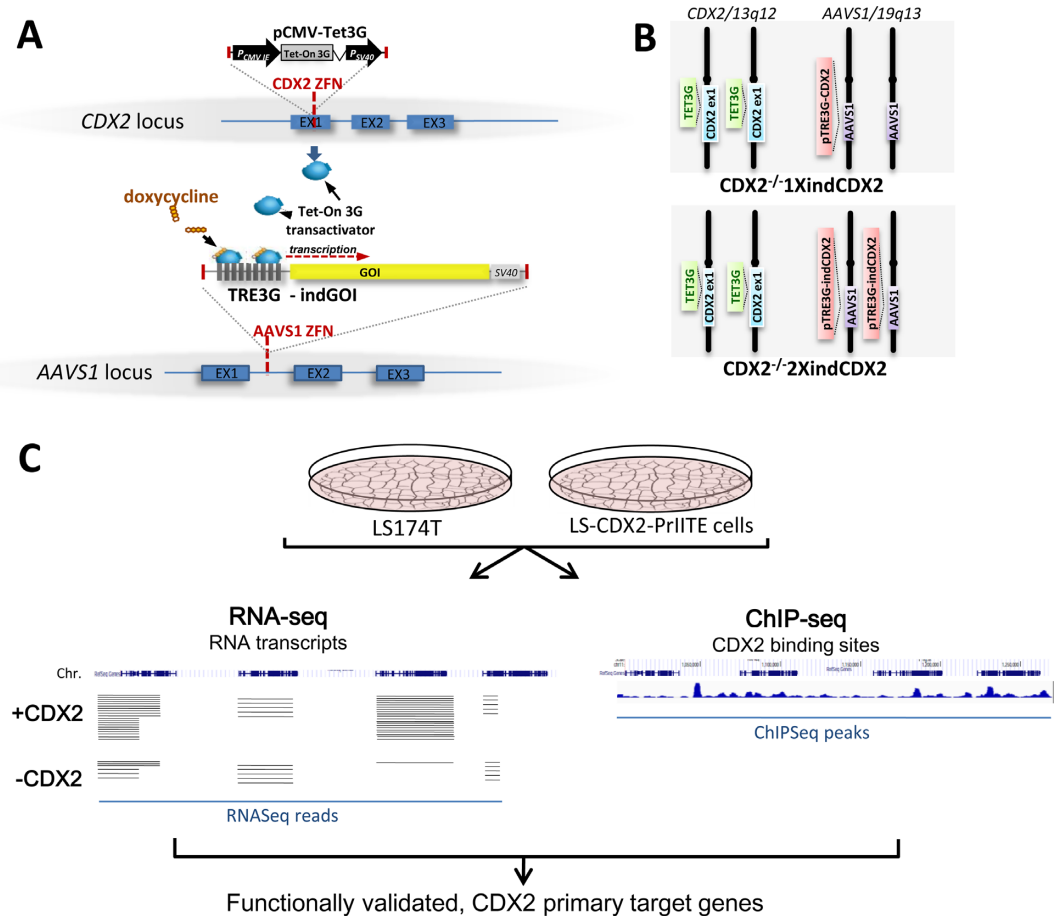

Figure 1. Overview of the established isogenic knockout-rescue model system based on precise integration of inducible elements (PrIITE) in LS174T. (**A**) pCDX2-pCMV-TET3G-ObLiGaRe donor vector (pCMV-Tet3G) was targeted to CDX2 exon1 by ZFN mediated targeted integration followed by pTRE3G-inducible Gene Of Interest (GOI) ObLiGaRe donor vector (pTRE3G-indGOI) targeting to the safe harbor AAVS1 locus. (**B**) Schematic illustration of the isogenic inducible CDX2 LS174T cell model system based on the mono- (upper panel) or bi- (lower panel) allelic precise genome targeting of CDX2 inducible elements, CDX2$^{-/-}$1XindCDX2 and CDX2$^{-/-}$2XindCDX2 respectively. (**C**) Illustration of the experimental workflow including deep transcriptomic analysis (RNA-seq) of wild type LS174T and isogenic indCDX2-PrIITE cells exposed to variable concentrations of doxycycline inducer and CDX2 target validation by genome-wide CDX2 ChIP-seq (ChIP-seq) analysis.

ure S11 for vector constructs used and generated. All plasmids were Sanger sequencing verified.

Cell culture and transfection

LS174T (ATCC-CL-188) human intestinal cell line was maintained in Ham's nutrient mixture F12/Dulbecco's modified Eagle's medium (1:1) supplemented with 10% fetal bovine serum and 1% glutamine. Different target 'knock in' strategies were undertaken, including use of 800 bp homology arms flanking the donor integration cassette D (data not shown). Successful LS174T targeted KI was only obtained when the ObLiGaRe strategy for improved target specific 'knock in' integration of donor constructs was employed (30). In brief, the ObLiGaRe strategy is based on use of existing ZFN binding elements that in an 'inverted' orientation flank the donor DNA construct. By co-nucleofection of ZFN's and donor DNA, double stranded breaks at both the specific chromosomal target site and flanking the donor plasmid will occur allowing cellular repair pathway mediated target specific integration of the linearized donor at the desired target site.

For stable *pCDX2-pCMV-TET3G-ObLiGaRe* integration the first exon of the *CDX2* gene was targeted through CDX2 CompoZr ZFNs driven integration (Sigma-Aldrich, USA) (AACTTCGTCAGCCCCccgcagTACCCGGACTACGGCGGTT, left and right ZFN binding sites capitalized and linker cloning site in lower case). 1×10^6 LS174T cells were transfected by nucleofection, simultaneously with the GFP/E2 Crimson tagged CDX2 ZFN plasmids (2 μg each) and the pCDX2-CMV-TET3G-donor vector (5 μg). Nucleofector solution T (Lonza, CH)/Nucleofector program T-020 were used for the electroporation procedure in an Amaxa Cell Line Nucleofector device (Lonza, CH). Cells were then cultured for 6 h at 37°C to stabilize, and then moved on to a 30°C cold shock. Two days after transfection, cells expressing both GFP and E2 Crimson were sorted out in a FACS ARIA III (BD BioSciences, USA) as previously described (29). The cell bulk was plated out to grow in collagen-coated plates and two rounds of cloning were then employed.

Single cell clones were screened by junction PCR using a primer flanking the 5′ *CDX2* target locus and a reverse oriented primer localized within the integrated *pCMV-TET3G* cassette (Supplementary Figure S1A and B). Based on IDAA assay (32), one of them (5B5) showed the presence of wt allele, representing heterozygozity for *p*CMV-TET3G integration at the *CDX2* locus (clone named $LS^{CDX2+/-}$), while the other two clones (3D6 and 5E2) were found to possess successfully integrated *p*CMV-TET3G transactivator at both alleles representing homozygozity for *p*CMV-TET3G (clones named $LS^{CDX2-/-}$#1 and $LS^{CDX2-/-}$#2 respectively).

1×10^6 cells from a *CDX2* KO clone containing mono or bi-allelic *p*CMV-TET3G transactivator inserted into the two *CDX2* alleles were then transfected by nucleofection with GFP/E2 Crimson tagged AAVS1 ZFN plasmids (2 μg each) and 5 μg *p*AAVS1-TRE3G-CDX2opt or *p*AAVS1-TRE3G-T2EYFP.

Clone characterization by polymerase chain reaction (PCR) and IDAA

In order to identify correct integration of *pCDX2*-CMV-TET3G, a region comprising a part of *CDX2* exon 1 upstream the ZFNs cutting site and left 800 bp homology arm and a part of the CMV promoter was amplified using Expand Long Template PCR System (Roche Applied Science, GE) (5′-CMV-TET3G junction PCR). Similarly, the identification of correctly integrated *p*AAVS1-TRE3G-CDX2opt or *p*AAVS1-TRE3G-T2EYFP fragments into *AAVS1* locus was performed by amplification of a region comprising a part of *AAVS1* upstream the ZFNs cutting site and a part of the *TRE* promoter (5′-TRE3G-CDX2opt or TRE3G-T2EYFP junction PCR). PCR reaction mixture consisted of 100 ng DNA, 2.5 μl buffer 1, 3.0 μl 1.25 mM dNTPs mix, 0.25 μl each primer at a concentration of 25 μM, 0.1 μl enzyme mix and water to a final volume of 25 μl. All junction PCR primers were obtained from TAG Copenhagen A/S, Denmark and are listed in Supplementary Table SIV. Amplification was done using the following touch down protocol. After preheating for 5 min at 95°C, 12 cycles were performed starting with denaturation for 45 s at 95°C, annealing for 15 s at 74°C with a decrease in annealing temperature of –1°C/cycle, and 2 min at 72°C, followed by an additional 25 cycles of 45 s at 95°C, 15 s at 64°C and 2 min at 72°C, followed by a final extension of 3 min at 72°C. PCR products were run in a 1.2% agarose gel, bands were gel purified and sequence confirmed by Sanger sequencing.

FACS-enriched stable clones (29) were screened by junction PCR using primers flanking the junction between the *CDX2* or *AAVS1* genes and the integrated cassette (Supplementary Figures S1 and S2).

Presence of unmodified CDX2 or AAVS1 target (WT allele presence test) was performed using the recently described IDAA method (32) and protocol guidelines (33). In brief, due to size of the donor constructs used (>2 kb, respectively) the respective CDX2 or AAVS1 ZFN target sites can only be successfully amplified if integration has not occurred at the respective target loci. PCR was performed using 100 ng DNA in 25 μl using AmpliTaq Gold (ABI/Life Technologies) for *CDX2* locus or TEMPase Hot Start DNA Polymerase (Ampliqon A/S, DK) for *AAVS1* locus using the recently described IDAA/tri-primer amplification conditions comprising primers flanking the respective target site in combination with a universal 6-FAM 5′-labeled primer (FamF), specific for a 5′-overhang attached to the forward primer. IDAA primers used were purchased from TAG Copenhagen A/S and are listed in Supplementary Table SIV. Fluorescently labeled amplicons were then analyzed by capillary electrophoresis based fragment analysis using an ABI3030 instrument (Applied Biosystems/Life Technologies, USA). Raw data obtained was analyzed using Peak Scanner Software V1.0 (Applied Biosystems/Life Technologies, USA).

Induction with Dox

The Tet-On3G system was induced with Dox (Sigma-Aldrich, USA) in a range between 0.004 and 4 μg/ml for CDX2 knock-in (KI) clones and 0.001–5 μg/ml for T2EYFP KI clones. Dox was added to the medium every 24 h for one or 2 days, and medium was exchanged every 48 h after Dox removal. Controls where Dox was not added to the medium were used (0 μg/ml of Dox).

Western blot of cell lysates

Whole-cell extracts were obtained by resuspension of cell pellets in RIPA buffer (50 mM Tris–HCl pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 0.1% sodium dodecyl sulphate) in the presence of complete protease inhibitors cocktail (Roche Applied Science, GE). Quantification of total protein was determined by bicinchoninic acid protein assay (Thermo Scientific™ Pierce, USA). 15 μg of protein extracts were then analyzed by standard SDS-PAGE, transferred to a nitrocellulose membrane (GE Healthcare Life Sciences, USA) and blotted on at 4°C with mouse monoclonal primary antibodies CDX2-88 1:500 (Biogenex, USA), anti-TetR 1:1000 (Clontech/Takara, USA), anti-MUC2, undiluted hybridoma supernatant and anti-actin 1:8000 (Santa Cruz Biotechnology, USA) in 5% BSA in PBS. Membranes designed for MUC2 blotting were previously treated with neuraminidase from *Clostridium perfringens* type VI (Sigma-Aldrich, USA) diluted in PBS to a final concentration of 0.2 U/ml, for 1 h at 37°C. Peroxidase-conjugated secondary antibodies (goat polyclonal anti-mouse-HRP 1:2000 for CDX2, TetR and MUC2 and goat polyclonal anti-rabbit-HRP 1:2000 for actin (Santa Cruz Biotechnology, USA) were used and developed with the ECL detection kit (Bio-Rad Laboratories, USA).

RNA extraction

Total RNA was extracted using RNeasy kit (Qiagen, GE) as recommended by the manufacturer.

Transcriptomics (RNA-Seq)

Polyadenylated RNA was isolated from total RNA using standard protocols (Dynabeads mRNA Direct Micro Kit, Ambion/Life Technologies, USA) starting with 5 μg total RNA. Library preparation for NGS using Ion Torrent technology (Life Technologies, USA) was carried out according

to the manufacturers recommendations (Ion Total RNA-Seq Kit v2, Ion torrent/Life Technologies, USA), and sequenced using an Ion Proton system (Life Technologies, USA). Quality control, quantification of RNA and libraries was carried out using using Agilent RNA 6000 Nano Kit or Agilent High Sensitivity DNA kit and Agilent Bioanalyzer (Agilent Technologies, USA). Sequencing reads were mapped to hg19 and bioinformatics analysis was conducted using CLCs Genomic Workbench (CLC bio/Qiagen, DK), STARBowtie2 followed by Cufflinks for expression analyses. In brief, ∼50 ng polyA RNA was fragmented down to 100–300 base fragments using RNaseIII for 1–3 min followed by adapter ligation, amplification for 9–14 cycles and barcoding using Ion Express RNA-Seq Barcode kit (Ion Torrent/Life Technologies, USA). The final library fragment size and concentration was determined by Agilent Bioanalyzer analysis followed by template preparation using Ion PI Template OT2 200 Kit v3 (Life Technologies, USA) and Ion One Touch System followed by NGS on an Ion Proton system using Ion PI™ Chip Kit v2. In general, two transcriptome libraries were barcoded and analyzed on one Ion PI v2 chip. Sequencing depth for the RNA-Seq data sets ranged from 34.5 to 44.6 mill reads/sample with 95.4–97.7% mappable reads to the hg19 reference data base. Mean read length for transcriptome RNA-Seq data ranged from 116 to 136 bp.

Genome-wide analysis of CDX2-binding sites by ChIP-seq.

LS174T cells grown for five days in a 30 × 30 cm culture dish were cross-linking and sonicated as described previously to generate fragments of ∼0.2 to 1.2 kb. Hereafter, the ChIP-protocol was performed as previously reported (34). Briefly, immunoprecipitation was done in four replicates and performed overnight at 4°C using specific antibodies to human CDX2 (CDX2-88, Biogenex, USA) and the influenza hemagglutinin (HA) epitope (rabbit polyclonal α-HA; Santa Cruz Biotechnology Inc, Heidelberg, Germany), used as a negative control. Immunocomplexes were recovered with 50 μl protein A/G beads (Invitrogen/ThermoFisher Scientific, USA). Verification of the enrichment in the CDX2-immunoprecipitated DNA samples was done by qPCR analyzing the DNA level from known CDX2-targets in the CDX2-immunoprecipitated samples and comparison it to level in the negative control (HA-immunoprecipitated DNA).

In order to increase the depth of the analysis, the ChIP library established was deep sequenced using Ion Proton instrumentation generating an approximate ≈54 million reads depth. Library preparation for NGS was carried out with the Ion Xpress Plus gDNA Fragment Library Kit according to User Bulletin 4473623: Ion Chipseq Library Preparation on the Ion Proton System (Life Technologies, USA), starting with 10 ng DNA. Quality control of the libraries was carried out using using Experion DNA 1K analysis kit and the Experion System (Bio-Rad Laboratories, USA). In brief, 10 ng DNA was end-repaired and ligated to barcode adaptors from the Ion Xpress Barcode Adapters 1–16 Kit (Life Technologies, USA), followed by nick repair and amplification for 18 cycles. The amplified libraries were subjected to two rounds of bead capture with the Agencourt AMPure XP Kit (Beckman Coulter, USA) to size-select fragments ∼160–340 bp in length. The final library concentration was determined by Qubit analysis (ThermoFisher Scientific, USA). The barcoded ChIP- and IP control-DNA-libraries were handled similarly to the RNA-Seq library and analyzed on a single Ion PI™ Chip v2. For the ChIP- and IP control-DNA-libraries a total of 54.5 mill reads (21.8 and 32.7 mill respectively) were obtained and 98.5% were mappable to the reference hg19 data base with a mean read length of 130 bp. The detection of CDX2 ChIP seq peaks and the location of the closest gene was performed using CisGenome version 2 (35).

Immunodetection by fluorescence-activated cell-scanning (FACS) analysis

Cells were trypsinized at different time-points after induction with Dox or after Dox removal and washed twice with PBS. After being centrifuged at 1200 rpm for 7 min, CDX2 KI cells were fixed in 4% PFA for 20 min., washed in PBS and permeabilized in cold methanol for 15 min. After washing with PBS, samples were incubated with the primary antibody (CDX2-88 clone, BioGenex, 1:500.) for 1 h at 37°C. Cells were then stained with FITC-conjugated AfinityPure immunoglobulin antimouse IgG (Jackson ImmunoResearch Laboratories, USA) diluted 1:100 in 0.05% BSA in PBS and then subjected to FACS in a FACS ARIA III (BD BioSciences, USA). T2EYFP KI cells were immediately scanned by FACS after cell trypsinization.

Immunocytochemistry

Cell slides were fixed in cold acetone for 15 min. For TetR anti-TetR cells were incubated overnight at 4°C with the primary antibodies (9G9 clone, Clontech/Takara, USA). Negative controls were performed by omission of primary antibodies. After washing, a rabbit anti-mouse Ig FITC-labeled secondary antibody (Dako A/S, DK) diluted 1:100 in PBS with 5% BSA was added for 45 min, protected from light. DAPI was used as a nuclear counterstain and slides were mounted in Vectashield mounting media (Vector Labs, USA).

For EYFP visualization on the induced T2EYFP KI clones, trypsinized cells were fixed in cold acetone for 10 min, DAPI was added and slides mounted. Samples were examined under a Zeiss fluorescence microscope equipped with DAPI and FITC interference filters. Images were acquired using a Zeiss Axioskop 2 and an AxioCam MR3 camera and Zeiss Application Suit software.

Real time *in vivo* video material

For the time lapse video, 1×10^6 CDX2$^{-/-}$2XindT2EYFP cells were seeded in a well of a six-well plate and 0.5 μg/ml of Dox were added after 24 h. Time lapse images started to be acquired under dark field immediately after Dox addition and every 20 min during 24 h using a Leica DMI 6000 timelapse microscope equipped with FITC interference filter and with CO_2 supply.

RESULTS

Precise integration of inducible transcriptional elements (PrIITE)

CDX2 knock-out by targeted integration of Tet3G transactivator elements. In order to eliminate the adverse cellular effects seen from prolonged exposure to Dox (25,26), we designed a strategy combining stable and controlled TeT-On transactivator/TeT3G expression with disruption of a target gene. First we established a Tet3G expressing LS174T cell line, in which the TET3G elements were targeted to *CDX2* exon1 CDS, thereby abrogating the endogenous gene function. Targeting *CDX2*, a master differentiation transcription factor (36), allowed us to determine the efficacy of the system in an unprecedented CDX2 knock out cell system (Figure 1). Multiple attempts at homologous recombination driven donor integration at the CDX2 exon 1 locus based on CDX2 ZFNs and a TET3G donor template flanked by 700bp homology arms did not give rise to any correctly targeted clones, likely due to low homologous repair capacity of LS174T cells used in this study (data not shown). *pCDX2-pCMV-TET3G-ObLiGaRe* donor vector (Figure 1A and Supplementary Figure S1A) was co-transfected together with *CDX2* CompoZr ZFN plasmids into LS174T cells, whereby *pCMV-TET3G transactivator* donor vector was specifically integrated into exon1 of *CDX2*, as illustrated in Figure 1A and Supplementary Figure S1A and B, by ObLiGaRe driven recombination (see Material and Methods section for details). Three clones (5E2, 3D6 and 5B5) were shown to include correct integration of the *p*CMV-TET3G construct and constitutively express the Tet3G transactivator (Supplementary Figure S1C), one clone, 5B5, was mono allelically targeted (designated $LS^{CDX2+/-}$) and CDX2 protein expression maintained at levels similar to the parental LS174T wt cell line and in the bi-allelic targeted *CDX2* knock-out (KO) clones, 3D6 ($LS^{CDX2-/-}$#1) and, 5E2 ($LS^{CDX2-/-}$#2), *CDX2* expression was completely abrogated (Supplementary Figure S1C and D). Constitutive Tet3G expression and abrogation of CDX2 was maintained after prolonged cell passaging and after freeze/thawing lack of CDX2 was correlated with significantly decreased protein levels of a well-know CDX2 target gene, *MUC2* (37) (Supplementary Figure S1C and D). Lastly, we tested the functionality of the mono and bi–allelically targeted *CDX2* KO cells by cellular transfection with known CDX2 enhancer reporter constructs and could show dose dependent decreased expression of a *HNF4A* reporter and >10-fold reduced expression of a *HEPH1* reporter (38) in the *CDX2* KO cells (Supplementary Figure S3).

We thus show, that *CDX2* KO by ZFN-mediated site specific *p*CMV-TET3G integration in LS174T cells was effectively accomplished and that stable Tet3G expression was maintained over time.

Safe harbor targeted integration of inducible T2EYFP elements. Having shown that both $LS^{CDX2-/-}$#1 and #2 display similar transactivator expression levels, an inducible T2EYFP model system was established in $LS^{CDX2-/-}$#1. The inducible *p*TRE3G-T2EYFP-*ObLiGaRe* donor vector was targeted to the safe harbor *AAVS1* locus by co-transfection with *AAVS1* CompoZr ZFN's into the $LS^{CDX2-/-}$#1 cells as illustrated in Figure 2A. Multiple correctly targeted indT2EYFP clones were obtained and two of these selected for further detailed analysis, one mono-allelic T2EYFP targeted clone 7E9 (hence forward referred to as $CDX2^{-/-}$1XindT2EYFP) and another bi-allelic targeted clone 5G8 (henceforward referred to as $CDX2^{-/-}$2XindT2EYFP), Supplementary Figure S2. Both clones maintained constitutive Tet3G expression and absence of CDX2 (Figure 2B).

LS174T indT2EYFP PrIITE cells display no leakiness and Dox induction is fast and reversible.. We first aimed to determine the leakiness of the inducible T2EYFP reporter in the $CDX2^{-/-}$1xindT2EYFP and $CDX2^{-/-}$2xindT2EYFP clones by FACS and immunofluorescence, which showed undetectable fluorescence in uninduced PrIITE cells (Figure 2C, D and E). Upon induction, immunofluorescence displayed the expected peri-nuclear reactivity, consistent with the correct sub-cellular localization of T2EYFP fusion protein in the Golgi apparatus (39) (Figure 2E).

We next determined the induction dynamics over time by *in vivo* live imaging, demonstrating that the induction dynamics is very rapid. T2EYFP expression is detectable within 4 h post induction and all cells display full expression of the reporter after 8–10 h (Supplementary video material). Importantly, the induction dynamics is fully synchronized and all cells simultaneously become positive for the T2EYFP reporter. To elucidate how Dox dosage might influence on the dynamics of the system, induction was performed using variable Dox concentrations ranging from 0 to 5 μg/ml. As shown in Figure 2C, modest induction was observed with dox concentrations <5 ng/ml, whereas significant induction required Dox concentration >10 ng/ml. Interestingly, these results suggest, that the induction potential is related to the number of inducible elements integrated into the system. For instance, using a 10 ng/ml Dox dosage rendered >90% of the $CDX2^{-/-}$2XindT2EYFP cells EYFP reporter positive, whereas <50% $CDX2^{-/-}$1xindT2EYFP cells were found positive for this dosage.

Next, the reversibility of the PrIITE cell system was quantified by FACS analysis of the EYFP expressing cell population after Dox removal from the media (Figure 2D). As observed, induction was fully reversible and after 10 days post Dox removal, all cells in the population returned to non-expressing levels. However, the dynamics in reversibility was found to be dependent on both number of inducible elements integrated and the Dox dosage used. Clearly reversal to non-expressing levels was achieved rapidly and within 5 days for $CDX2^{-/-}$1XindT2EYFP and slower for $CDX2^{-/-}$2xindT2EYFP using the lower 50 ng/ml induction dosage.

We thus demonstrate that the PrIITE cell system displays no leakiness, is fully reversible, and that the induction dynamics occurs fast and in a synchronized manner and require below cellular stress inducing Dox concentrations.

LS174T indCDX2 PrIITE cells show no leakiness and ectopic CDX2 expression is tightly controlled requiring minimal Dox induction levels. The role of CDX2 in intesti-

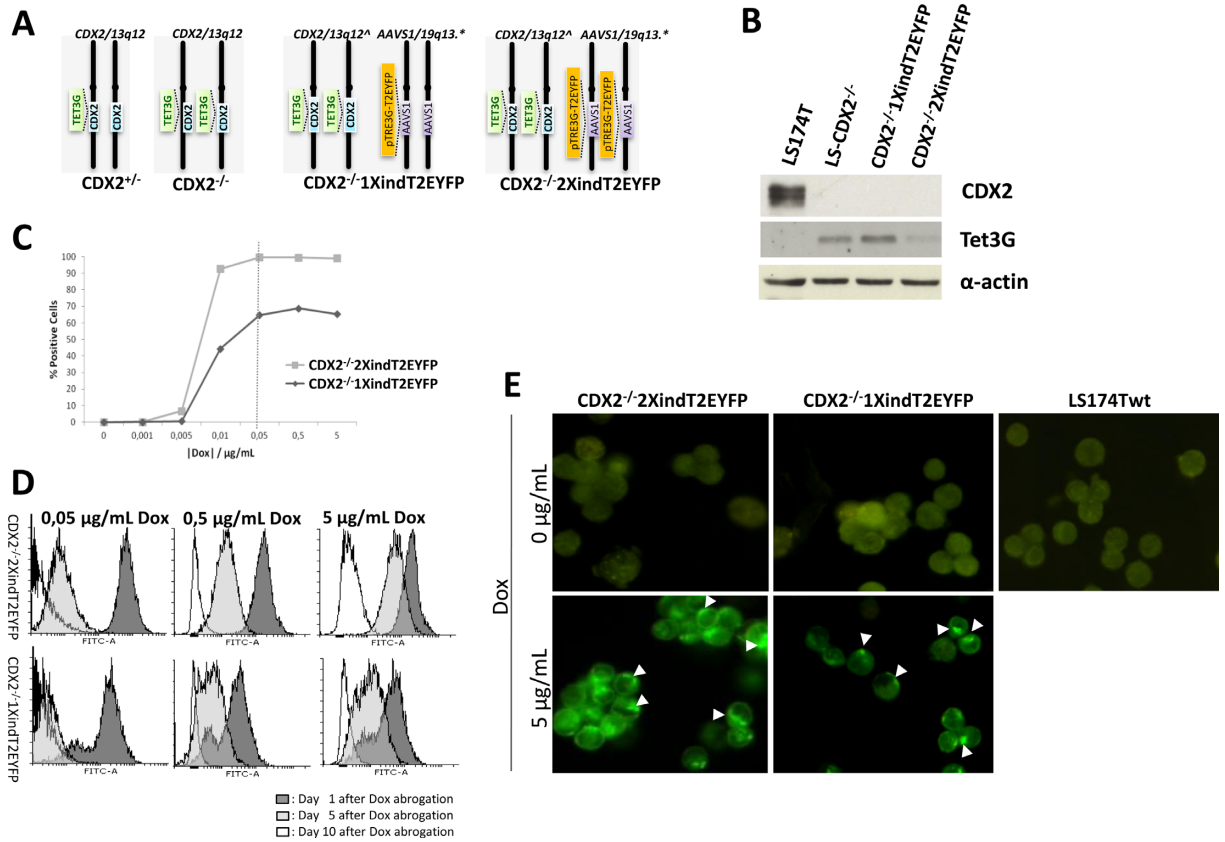

Figure 2. Overview of the proof of principle system using inducible T2EYFP reporter in isogenic LS174T-PrIITE cells. (**A**) First pCMV-Tet3G was targeted to *CDX2* exon1 of LS174T cells generating mono- or bi-allelic CDX2 KO cells, $CDX2^{+/-}$ and $CDX2^{-/-}$ respectively. Next, cells were re-targeted with *pTRE3G-T2EYFP ObLiGaRe-donor vector* (pTRE3G-T2EYFP) directed to the safe harbor *AAVS1* locus generating $CDX2^{-/-}$1XindT2EYFP and $CDX2^{-/-}$2XindT2EYFP cells. (**B**) Western blot analysis of LS174wt, $CDX2^{-/-}$, $CDX2^{-/-}$1XindT2EYFP and $CDX2^{-/-}$2XindT2EYFP cells show absence of CDX2 and stable TeT3G expression. (**C**) FACS based determination of percentage indT2EYFP positive cells after induction with different Dox concentrations. The graph shows that induction is dose and copy number dependent in $CDX2^{-/-}$1XindT2EYFP and $CDX2^{-/-}$2XindT2EYFP PrIITE cells. Approximate 100% T2EYFP positive cells was achieved using >0.005 µg/ml Dox concentration for $CDX2^{-/-}$2XindT2EYFP. Maximum 70% positivity was achieved regardless of the higher Dox induction concentrations used for $CDX2^{-/-}$1XindT2EYFP cells. Importantly, $CDX2^{-/-}$2XindT2EYFP cells show complete absence in T2EYFP fluorescence in uninduced cells thus, no leakiness of T2EYFP LS-PrIITE cells was detectable. Critical dox concentration above which adverse cellular stress (24–26) is caused is indicated by a dotted gray line. (**D**) Reversibility of induction as determined by FACS analysis of fixed cells 1, 5 or 10 days post 48 h induction with variable Dox induction concentrations. Notably, reversibility was copy number and dox dependent and full reversibility within 10 days was only achieved for $CDX2^{-/-}$1XindT2EYFP PrIITE cells using the lower 0.05 µg/ml Dox concentration. (**E**) Sub cellular localization of inducible T2EYFP as determined by fluorescence of fixed trypsinized indCDX2 PrIITE cells with or without dox induction. Arrow heads indicate expected Golgi localization of T2EYFP.

nal gene regulation is well known. However, to the best of our knowledge no studies so far described the effect of complete ablation of *CDX2* from the cellular genome. In a single study *CDX2* has been suggested as a lineage survival oncogene when amplified in colorectal cancer (40). In order to distinguish endogenous from ectopically expressed *CDX2* transcripts the inducible *CDX2* open reading frame was codon optimized (hereafter referred to as indCDX2). *pTRE3G-indCDX2-ObLiGaRe* donor vector was targeted to the *AAVS1* safe harbor site of $LS^{CDX2-/-}$#1 cells generating clones 7D9 ($CDX2^{-/-}$1XindCDX2) and 6D6 ($CDX2^{-/-}$2XindCDX2) possessing one or two inducible codon optimized CDX2 copies respectively (Figure 1B and Supplementary Figure S4). Constitutive Tet3G expression was maintained in targeted cells and $CDX2^{-/-}$2XindCDX2 or $CDX2^{-/-}$1XindCDX2 cell induction dynamics over 24 h were similar (Figure 3A and B). To test the reversibility of the CDX2 PrIITE system built, we induced with a range of Dox concentrations (0–4 µg/ml) for 48 h followed by removal of Dox from the medium. As expected, no expression of indCDX2 was revealed in non-induced cells followed by fast dose dependent Dox induction dynamics (Figure 3C). Importantly, for the lowest concentration of Dox used (0.004 µg/ml), the induction dynamics for the $CDX2^{-/-}$ 1XindCDX2 clone was slower than for $CDX2^{-/-}$2XindCDX2 (Figure 3C). Of notice, the indCDX2 protein levels, reached for $CDX2^{-/-}$1XindCDX2 with 4 ng/ml Dox dose, were similar to the endogenous CDX2 levels detected in LS174Twt cells (Supplementary Figure S4E) while for $CDX2^{-/-}$2XindCDX2 induction reached 120× higher levels relative to LS174Twt levels (Supplementary Figure S4E). In contrast to the dynamics in reversibility for indT2EYFP KI clones (Figure 2), reversibility of indCDX2 induction appeared faster and required very modest Dox induction dosages in the sub cell stress inducing nanogram/ml range.

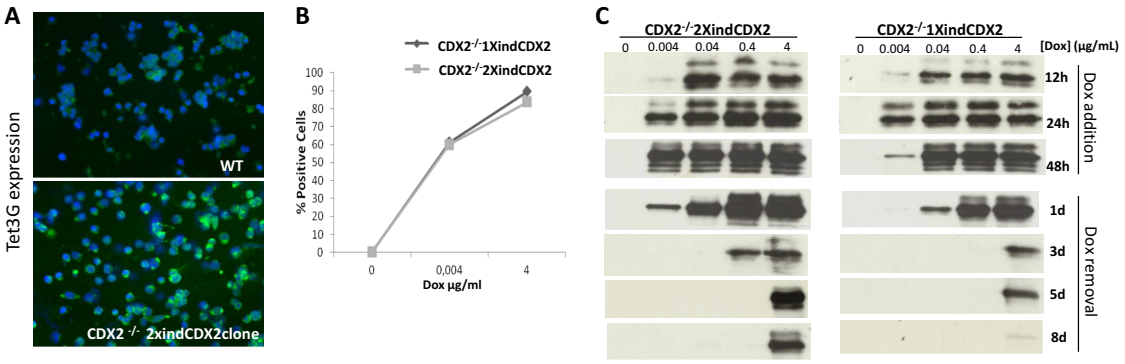

Figure 3. Induction dynamics of isogenic inducible CDX2 LS174T PrIITE cells. (**A**) Anti-TeT3G transactivator immunofluorescence in CDX2$^{-/-}$2XindCDX2 cells and not in LS174T wt cells. (**B**) 24 h induction kinetics of indCDX2 using different Dox concentrations. >60% of CDX2$^{-/-}$1XindCDX2 and CDX2$^{-/-}$2Xind CDX2 cells become indCDX2 positive after 24 h 0.004 μg/ml Dox induction. Use of higher Dox concentrations higher percentage of cells become CDX2 positive after 24 h Dox treatment. (**C**) Dynamics in reversibility of the PrIITE system using different Dox induction concentrations. CDX2$^{-/-}$2Xind CDX2 (left panel) and CDX2$^{-/-}$1Xind CDX2 (right panel) cells were induced 48 h with the Dox concentrations indicated above the panels. Hereafter Dox was removed and post induction indCDX2 protein levels were determined by Western blot analysis 1, 3, 5 or 8 days post Dox removal. Maximum indCDX2 induction levels were achieved using 0.004 μg/ml Dox concentration for CDX2$^{-/-}$2XindCDX2 whereas 0.040 μg/ml Dox was required for CDX2$^{-/-}$1Xind CDX2.

Based on western blot analysis indCDX2 protein was undetectable within 3 days post Dox removal for both mono- and bi-allelic clones (Figure 3C).

Validating the tightness in inducibility of the PrIITE by deep transcriptome analysis. To confirm the tightness of the PrIITE system and in order to determine the CDX2 downstream target genes, LS174Twt, uninduced and induced CDX2$^{-/-}$2XindCDX2 or CDX2$^{-/-}$1XindCDX2 cells were RNA-seq transcriptome profiled (Figure 4A, and Supplementary Figure S5). To ensure that the isogenic PrIITE cells generated were representative of the mosaic moderately well differentiated LS174Twt cell population (41), all comparative analysis here and in the following sections included LS174Twt cells. Approximately 10 000 genes with RPKM >1 were shown to be expressed in the wt cells and the expression profiles for the mucin genes *MUC2* and *MUC5AC* correspond to the profiles previously reported for LS174Twt cells with high expression for the former and low expression for the latter (42,43) (Figure 5B). CDX2 target genes were defined by the following criteria; genes with RPKM >1 in wt cells where RPKM values in uninduced PrIITE cells (*CDX2*$^{-/-}$) were reduced >4× (4×↓) and increased >2× (2×↑) after re-induction relatively to the uninduced RPKM values. In depth analysis of the RNASeq reads that mapped to the human exon1 *CDX2* gene locus, including the region across the *CDX2* ZFNs cutting site, confirmed the presence of out of frame transcripts possessing deletions in CDX2$^{-/-}$2XindCDX2 exon1, while the heterozygous CDX2$^{-/-}$1XindCDX2 clone expressed intact endogenous *CDX2* transcripts present at ≈40% of the LS174Twt transcript levels (Supplementary Figure S5A). Analysis of the unmapped RNA-seq reads confirmed the presence of the Tet3G transactivator transcripts in CDX2$^{-/-}$ cells (Supplementary Figure S5B) and importantly, no indCDX2 transcript was detected in the uninduced state of the CDX2$^{-/-}$2XindCDX2 clone (Supplementary Figure S5B), thus confirming the biochemical and immunofluorescence results obtained previously.

Taken together, by biochemical analysis, immunofluorescence and deep transcriptome analysis, we have shown that the PrIITE system allows for absolute control of inducible and reversible gene expression of CDX2 and it's downstream target genes.

Validating the transcriptome identified CDX2 target genes by genome-wide ChIP-seq analysis. Reasoning that the RNA-seq identified CDX2 target genes could be direct or indirect targets of CDX2, we wanted to validate the RNA-seq data set by genome wide chromatin immunoprecipitation sequence analysis (ChIP-seq). We thus used ChIP-seq to identify true CDX2 binding elements in the vicinity of the RNA-seq identified CDX2 target genes. ChIP was performed on LS174Twt cells using a well-known anti-CDX2 monoclonal antibody essentially using a previously reported procedure (38). A LS174T non-immunoprecipitated library was deep sequenced in parallel and used as ChIP-seq background control. An additional requirement was added to the RNA-seq criteria for positive scoring of primary CDX2 target genes, in that a ChIP-seq peak (CDX2 binding element) were to be present within or 1 kb up or downstream of the target gene. The combined RNA- and ChIP-seq criterion identified 31 direct CDX2 target genes (Figure 4B and Supplementary Table SI). Among these genes several known CDX2 regulated targets were found, such as *TFF3*, *MUC2* and *CDX2* itself, but several novel targets were also found, including *MUC5B, MUC5AC*, *MUC6*, *GPA33* and *LDLR* (Figure 5). Furthermore, *in vitro* promotor analysis confirmed the CDX2 dependent regulation of the enhancer identified in the novel CDX2 target gene *GPA33* (44) (Figure 6) and thus, substantiated both our RNA- and ChIP-seq findings. Interestingly, a significant intergenic CDX2 binding element was identified in the 300 Mb 11p15 region containing the mucin genes *MUC6*, *MUC2*, *MUC5B* and *MUC5AC* (Supplementary Figure S6A). Clearly our transcriptome results reveal a significant concerted down-regulation of this gene cluster and suggest CDX2 as a 'locus control gene' (Figure 5B). As expected, the non CDX2

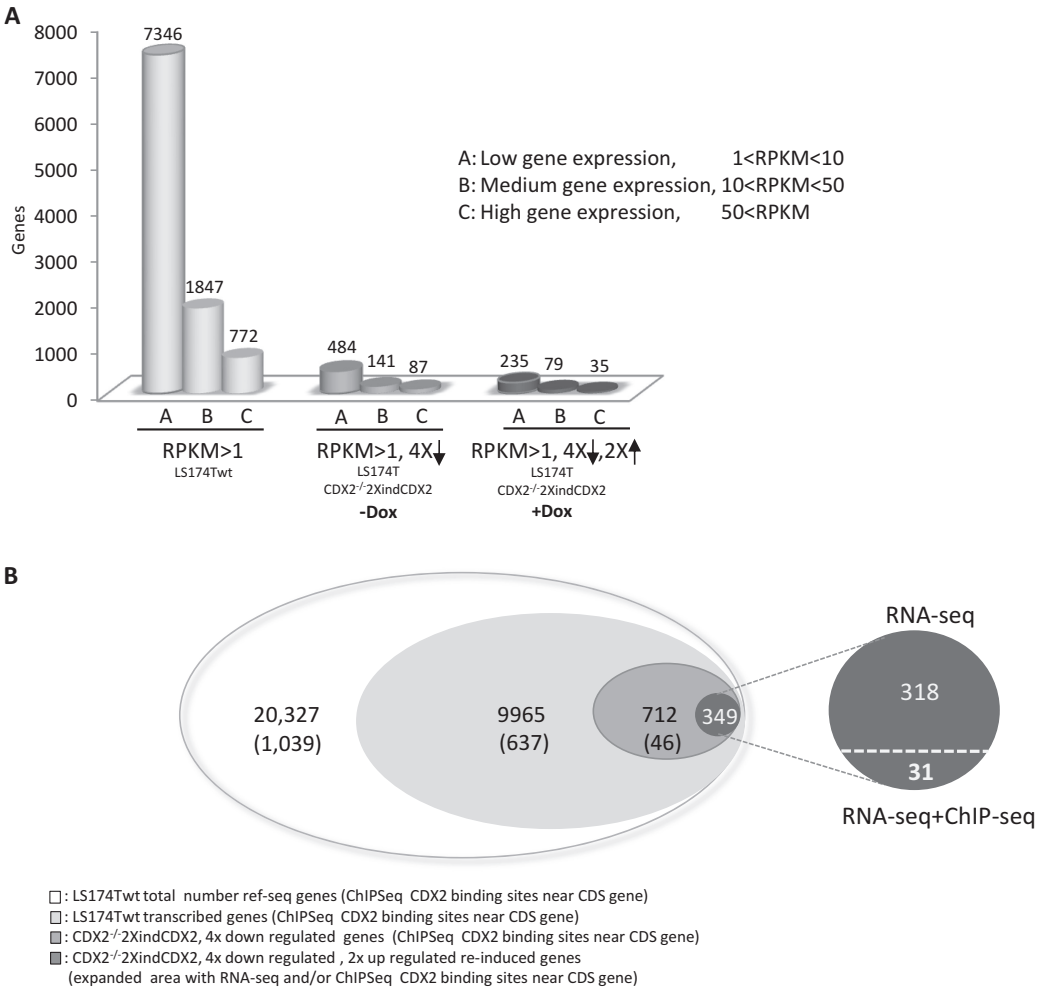

Figure 4. Deep RNA-seq and genome wide ChIP-seq analysis of LS174Twt and induced or uninduced CDX2$^{-/-}$2XindCDX2 PrIITE cells. (**A**) Of the 20 327 human genes included in the analysis around 9965 were found to be expressed with RPKM >1, the majority hereof (7346) were low expressed genes with 1 < RPKM < 10 (**A**), followed by medium expressed genes (1847) with 10 < RPKM < 50 (**B**) and highly expressed genes (772) with 50 < RPKM (**C**). In uninduced CDX2$^{-/-}$2XindCDX2 (−Dox) PrIITE cells (lacking CDX2), 712 genes expressed in LS174Twt were 4-fold down regulated and after indCDX2 re-induction (+Dox) 349 of these genes were 2-fold upregulated and thus suggested to be CDX2 downstream target genes. (**B**) In order to determine to what degree the genes identified represent direct CDX2 target genes, we included a LS174T ChIP-seq data set and seeked for the presence of potential CDX2 regulatory elements in the vicinity of the RNA-seq suggested CDX2 target genes. Inclusion of this criteria (numbers shown in parenthesis) in addition to the RNA-seq criterion defined above resulted in a dramatic decrease in potential direct CDX2 targets going from 349 to 31 genes. The expanded view at right depicts the number of targets identified using the RNA-seq or RNA-seq+ChIP-seq criteria. In the latter case the targets indicated below the stippled line are listed in Supplementary Table SI.

regulated mucin genes *MUC1* and *MUC7* did not possess any CDX2 binding elements and did not respond to CDX2 changes (Figure 5B and Supplementary Figure S6B). Previously, we have identified CDX2 as a regulator of glycosyltransferase encoded *ST6GALNAC1* expression (45). However, in this study, no CDX2 binding elements were identified in the *ST6GALNAC1* locus (Supplementary Figure S6B), which suggests that CDX2 regulates the expression of this gene by an indirect effect.

The results also show that several known CDX2 targets such as *FUT2* (46), *B3GALT5* (47), *ALP1* (48) and *SI* (28) in LSCDX2 PrIITE cells were unaffected by CDX2 changes (Figure 5B and Supplementary Figure S7) and notably, in the latter three cases these genes were shown not to be expressed in LS174Twt cells. Of notice, the only glycogenes (49) otherwise affected by CDX2 were *FUT3* and *GALNT1* which in the latter case responded inversely to CDX2 changes (Figure 5B).

In order to assess to what degree the global reversibility in gene expression for LSCDX2 PrIITE cells was retained, RPKM values from LS174Twt and LSCDX2 PrIITE cells exposed to different Dox concentrations ranging from 0 to 40 ng/ml were plotted and the regression line for each data calculated (Figure 7 and Supplementary Figure S8). The results show that global reversibility of CDX2 controlled gene expression was profound, and in many cases full reversibility was observed when the lowest 4 ng/ml Dox concentration was used with CDX2$^{-/-}$2XindCDX2 cells. Clearly, the combined RNA-seq and ChIPseq filtering strategy employed improved the specificity in defining true CDX2 regulated genes in our data set (Figure 7C). We speculate, that the observed highest normalization levels of

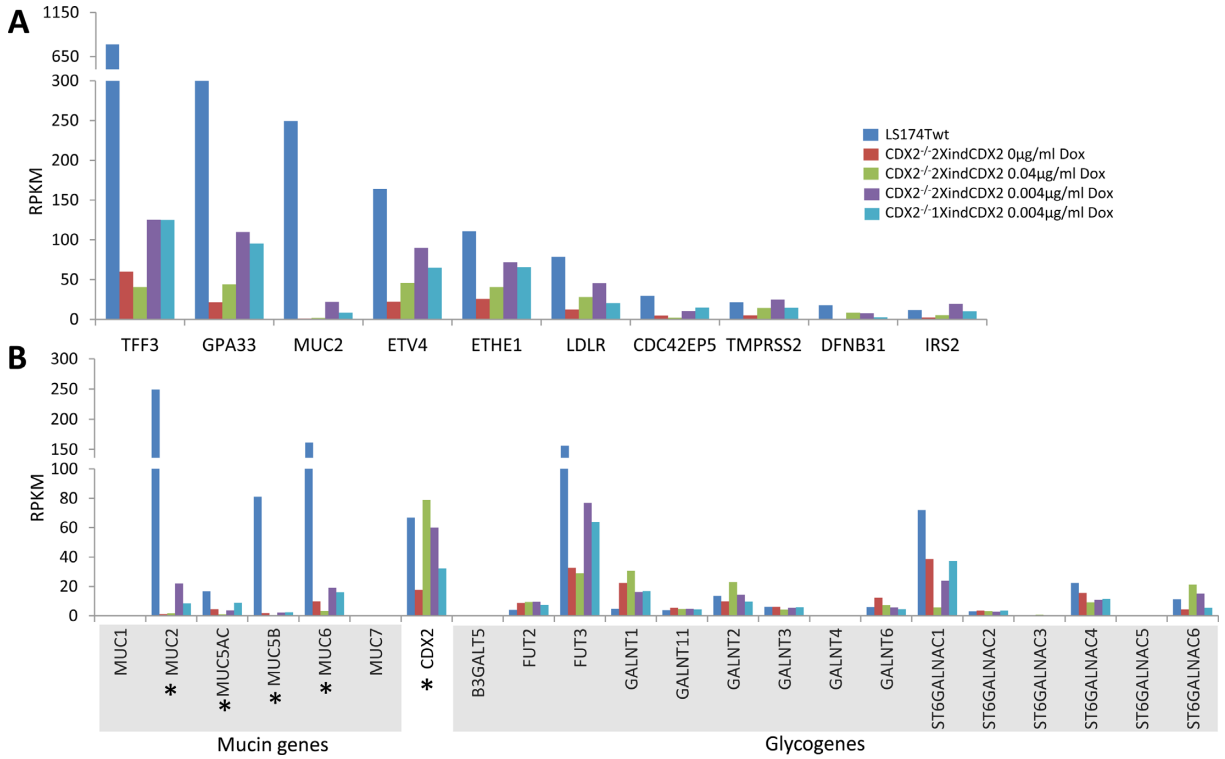

Figure 5. CDX2 regulated target genes. (**A**) Based on the included RNA-seq and ChIPseq criteria (see Figure 4 and Supplementary Table SI) top 10 primary targets downstream of CDX2 with most abundant transcripts (highest RPKM values) were ranked. Well known CDX2 target genes were identified such as *TFF3* and *MUC2*, but novel targets were also identified including *GPA33* and *LDLR*. (**B**) Careful analysis of 208 glycogenes (49) controlling the cellular glycophenotype and 23 mucin genes revealed that *MUC5B*, *MUC6* and *MUC5AC* were controlled by CDX2, whereas *MUC1* and *MUC7* did show no correlation with CDX2 expression. No CDX2 correlation with *FUT2* and *B3GALT5* could be demonstrated, which is in contrast with previous reports. On the other hand, *ST6GALNACI* and -*VI* were shown to correlate with CDX2 expression although no CDX2 binding element could be identified within the vicinity of the CDS of these genes. Interestingly, an inverse correlation was observed for *GALNT1* regulation similar to *SOX2* (see Supplementary Figure S7), with upregulation in the absence of CDX2 expression. Interestingly, the novel CDX2 target *GPA33* gene product has been shown to be post translationally modified by the *GALNT1* encoded enzyme. Genes with ChIP-seq identified CDX2 binding elements in the gene locus vicinity are indicated by an *.

re-induced CDX2 for 4 ng/ml Dox may be due to the role of CDX2 as a molecular rheostat potentially controlling: (i) self-renewal as indicated in Figure 5 and/or (ii) regulated expression of downstream target genes within discrete and narrow CDX2 expression limits (50). Lastly, in order to determine if the genetic manipulations had changed the differentiation state and stemness of CDX2 PrIITE cells generated, we determined the expression profile of 24 marker genes before and after genetic manipulation and/or CDX2 re-induction (Supplementary Figure S7). In all cases (except for *MUC6*), the gene expression profiles of genetically manipulated cells relative to LS174Twt was retained and displayed a distinct intestinal expression pattern. Interestingly, the stemness related gene, *SOX2*, is not expressed in the wt cells, but strongly expressed upon CDX2 KO and clearly decreased upon reinduction. This may suggest that also stemness regulation is CDX2 dependent. The CDX2 targets identified in our data-set is to some extent in agreement with previous findings in a related cellular background (38) (Supplementary Table SI). Of note, the karyotype of the LS174T cells used in our study was determined and found to be in complete agreement with the originally near normal published karyotype (41) (Supplementary Figure S9).

DISCUSSION

Gene regulatory mechanisms are complex and profit from the development of systems in which gene and protein expression is tightly controlled. Tet-regulated expression systems have been widely used for inducible protein expression in mammalian cells. However, current optimized Tet-on systems are still hampered by residual levels of gene expression in the uninduced state in both transiently and stably transfected cells (18,21,22). We hereby demonstrate that the PrIITE system resolves all issues related to leakiness and importantly requires use of tetracycline dosages below cell stress inducing concentrations (24). We first prove the PrIITE concept using a Golgi targeted EYFP model reporter (31). With this tightly controlled inducible reporter system at hand, we were able to demonstrate, to the best of our knowledge for the first time, that leakiness relates to increased expression of the TET3G transactivator element and not increased presence of the inducible GOI, in this case the T2EYFP reporter, Supplementary Figure S10. We furthermore show that stable random integrated pCMV-TET3G transactivator elements causes leaky expression in a substantial sub-population of the cells and we speculate that this sub-population of cells enable uncontrolled Dox induction levels beyond what is achievable in the tightly

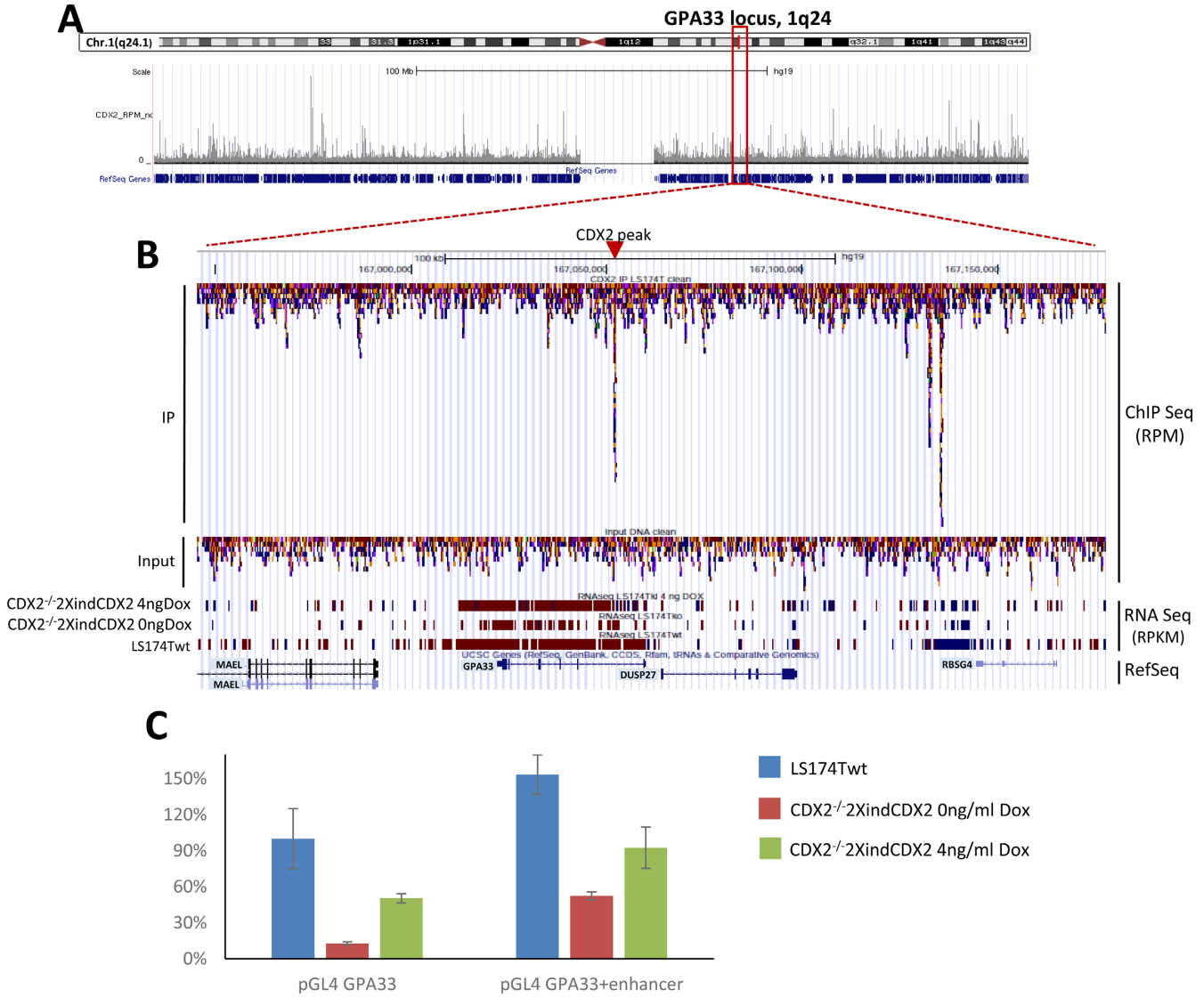

Figure 6. Identification and validation of the novel CDX2 target gene, *GPA33*. (**A**) Panel shows chromosome 1 coverage of CDX2 ChIPseq reads (CDX2IP) and control (input), calculated in 300 bp sliding windows as RPM (reads per million). For each window, input RPM was subtracted from CDX2IP RPM and normalized values were saved as a bedgraph file for visualization in the UCSC Genome Browser. Major CDX2 binding sites (peaks) are shown with indication of the *GPA33* locus framed in red box. (**B**) Expanded display of the GPA33 locus, showing detailed identification of CDX2 ChIPSeq peaks (ChIPseq/RPM) in both LS174T CDX2IP and control input samples. Major *GPA33* CDX2 binding element is indicated by red triangle (CDX2 peak). RNA-seq panel displays the density of individual reads in CDX2$^{-/-}$2XindCDX2 with or without Dox induction (4 or 0 ng respectively) and LS174Twt cells. RefSeq gene panel displays the genomic organization of neighboring genes in the *GPA33* locus. (**C**) *In vitro* promoter analysis of the 0.6 kb CDX2 enhancer element identified in the *GPA33* intron. pGL4-GPA33 and pGL4-*GPA33*enhancer represent the reporter plasmid without or with the enhancer element included. Reporter expression is clearly dependent on both GPA33 CDX2 enhancer and CDX2 presence.

controlled PrIITE cells generated, Supplementary Figure S10B. We next exemplify the potential of the system by clarifying the role of CDX2 in intestinal cells as a consequence of controlled *CDX2* expression. This model system shows that CDX2 predominantly acts as a transcriptional activator and only a limited number of genes seemed to be repressed by CDX2 presence, SOX2 being an example hereof (Supplementary Table SII). We had anticipated that CDX2 targeting in LS174T cells could become problematic due to its proposed essential cellular function and pre-designed our targeting strategy for dealing with essential genes. This was accomplished by incorporating a 'landing pad' in the EPB71 AAVS1 donor integration vector, just downstream of the inducible gene of interest (Supplementary Figure S11). The 'landing pad' encodes the Safe Harbor #1 (SH1) sequence derived from the CHO genome, that has successfully been utilized by us (51) and others (52) for ZFN mediated target integration in CHO cells and thus represents a unique site when integrated in human cells devoid of this sequence. This allows for subsequent donor target integration into the SH1 site. Establishment of a PrIITE cell model for an essential gene can thus be accomplished, by targeting of both the TET3G transactivator and inducible pTRE3G codon optimized GOI transcriptional elements to

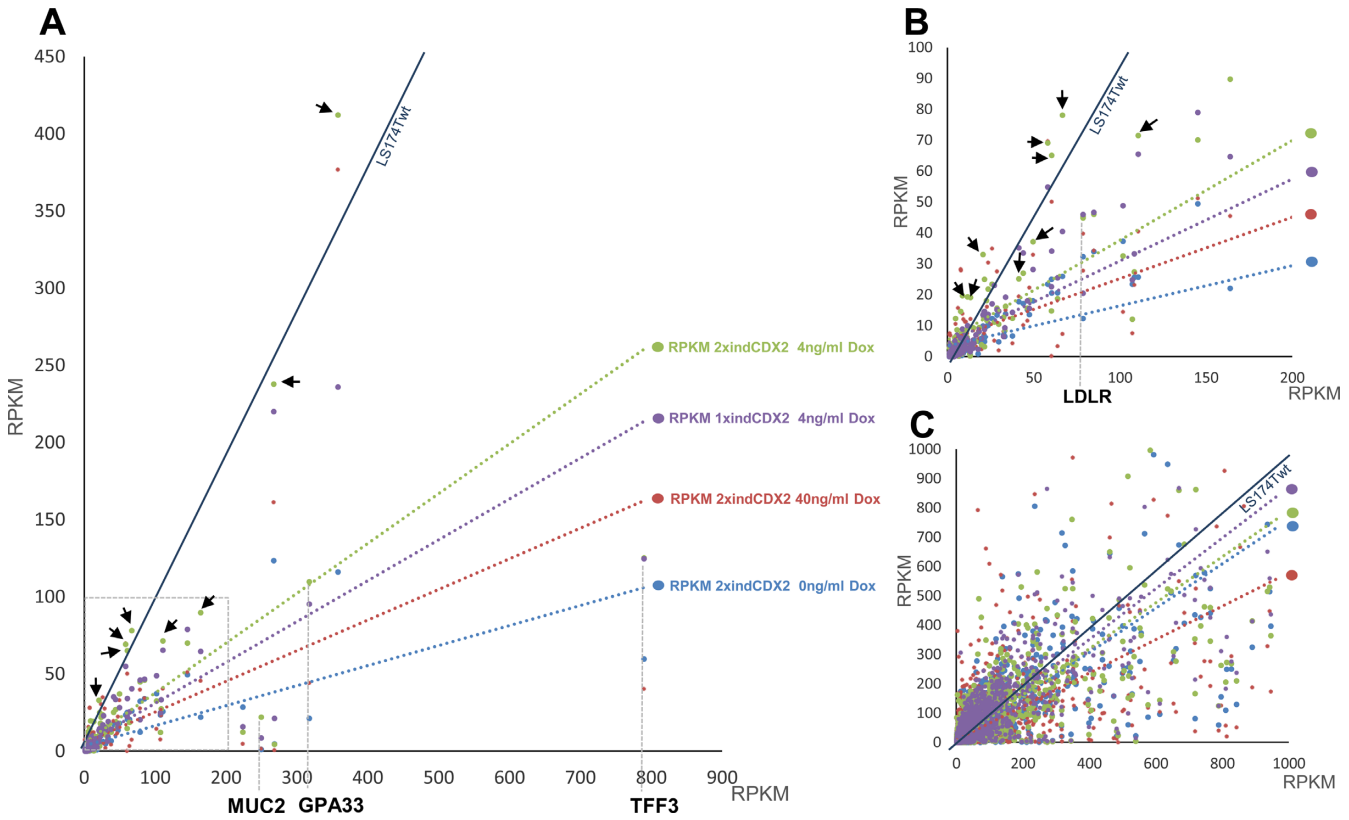

Figure 7. Global Dox induction reversibility of the PrIITE cells relative to LS174Twt cells. In order to include enough data points in the analysis, filtering criteria included all genes with RPKM >1, that were 2× reduced in CDX2$^{-/-}$ KO cells and had a CDX2 binding element/peak in the gene locus. RNA-seq data (RPKM) from wildtype LS174T was plotted against RPKM values from uninduced CDX2$^{-/-}$2XindCDX2 cells without CDX2 expression (blue dotted line, 2XindCDX 0 ng DOX) and under different conditions of CDX2/Dox induction (ng/ml) (green dotted line, 2XindCDX2 4 ng DOX; purple dotted line, 1XindCDX2 4 ng DOX, red dotted line, 2XindCDX2 40 ng DOX). The regression line for each data set is plotted. The dark blue solid line represents the expression levels in LS174Twt cells. (**A**) Global display of RNA-seq RPKM values for the samples analysed. RPKMs for the known CDX2 target genes *MUC2 and TFF3* and the novel target gene, *GPA33*, are indicated by grey dotted vertical lines. Box with gray dotted line is area shown in panel B. Representative examples of genes displaying full reversibility are indicated with arrow heads. (**B**) Zoom from dotted box from panel A with indication of RPKM values for the novel CDX2 target gene, *LDLR*. Representative genes for which complete reversibility was observed are indicated by arrow heads. (**C**) Global display of RNA-seq RPKMs for genes fulfilling the RPKM >1 and 2× reduction filtering criteria but without fulfilling the criteria of a CDX2 binding element/peak in the gene locus. Note the unordered dispersed display of RPKMs relative to the LS174Twt regression line, suggesting that a large proportion of the CDX2 target genes only based on RNA-seq data are not direct targets of CDX2.

the AAVS1 locus, ex. by AAVS1 targeting of the former gene followed by SH1 targeting of the later gene. Hereafter, precise target inactivation of the endogenous gene of interest in the presence of Dox would ensure continued expression of the inducible codon optimized version of the essential GOI. In this project, we did not take advantage of this option, since in the LS174T cellular model CDX2 has not proven to act as a lineage-survival oncogene in contrast to what is observed in other cell line models (40).

Our results also show that LS174T cells have a poor capacity to integrate homology arm flanked donor templates, which is likely attributed to the lacking HR capacity of these and other cell lines used (53–55). In general, precise donor integration remains a challenge in the field and in spite of the fact that CRISPR efficiently induces double stranded breaks, CRISPR mediated donor integration importantly comes at the expense of: (i) 3–20-fold higher indel formation rates at the non-targeted allele (56,57) and (ii) significant off-target/random integration frequencies (58). These concerns were evident in a recent *in vivo* study based on CRISPR target integration of inducible TRE3G promoter elements, where 'leaky' expression in 33% of clones analyzed was identified (59). In light of these CRISPR related issues, the ObLiGaRe targeting strategy by Marcello Maresca *et al.* remains to be an efficient and reliable target integration strategy for most cells (60–63), including HR compromised cells such as CHO cells (64,65) and colorectal cancer lines as described in this study and by Marcello Maresca et al.

A whole set of novel genes, including *GPA33* (44), *LDLR* (66), *MUC6*, *MUC5B* and *MUC5AC* were identified as CDX2 targets. To our knowledge, this has not previously been demonstrated and suggests *CDX2* as a locus control gene of the 11p15 mucin gene cluster (67–69) that includes *MUC2*, *MUC6*, *MUC5B* and *MUC5AC*. In this respect, a recent study of 295 gastric adenocarcinomas has shown that *GPA33*, *MUC6* and *MUC2* were among the 10 most abundant differentially regulated genes in 295 analyzed microsatellite instable (MSI) colorectal adenocarcinomas (70). Although *CDX2* profiling was not evident in this study it is stimulating to assume that MSI has impacted on *CDX2* expression in the tumors studied. Furthermore, *GPA33* is

expressed in cancer of the gastrointestinal tract, namely in over 95% of human colon cancers and has thus been suggested as an attractive novel colorectal carcinoma therapeutic target (71,72). Of particular notice, *GPA33* gene product has been shown to be post translationally modified by the *GALNT1* encoded GalNAc-transferase 1 enzyme (73,74), that we here show is inversely influenced by CDX2 (Figure 5B). Our finding that *LDLR* is a major direct CDX2 target gene is surprising since *LDLR* has been primarily found to be regulated by sterol regulated element-binding proteins (SREBPs) (75,76). We speculate that CDX2 plays novel roles in the co-regulated appearance of these novel gene targets.

In summary, we hereby describe a novel approach for generating an inducible 'non leaky' isogenic knockout-rescue system. The principle is based on precise genome integration of defined copies of third generation Tet-On elements, which enables for an 'all or none' context for studying the mechanism by which a given gene works in a cell. We exemplify the utility of the system by establishing a colonic CDX2 PrIITE cell system that uncovered novel molecular interactions governed by CDX2.

MATERIALS & CORRESPONDENCE

All reagents and cell lines used in the study are available upon request for research purposes under a material transfer agreement, except for TRE3G and TET3G containing plasmids, due to restrictions over distribution of plasmids containing the TRE3G promoter or TET3G ORF. The RNA-seq and ChIP-seq data shown has been uploaded at the NIH GEO server: http://www.ncbi.nlm.nih.gov/geo/ and is accessible using the following accession number GSE97273. Plasmids described are available from Addgene (https://www.addgene.org/). All correspondence should be addressed to Eric P. Bennett.

NOTE ADDED IN PROOF

Extended experiments conducted during proofs of this manuscript revealed that the GPA33 promoter described in Figure 6 possess bi-directional transcriptional activity, likely mediated through presence of multiple LINEs contained within the 1KBp GPA33 promoter analyzed. LINEs have been shown to mediate bi-directional expression (Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P. and Myers, R.M. (2004) An abundance of bidirectional promoters in the human genome. Genome Res.14:62–66; Core, L.J., Waterfall, J.J. and Core, J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science, 322:1845–1848) potentially affecting 4% of all human genes (Speek, M. (2001) Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol. Cell. Biol. 2001, 1973–1985; Criscione, S.W., Theodosakis, N., Micevic, G., Cornish, T.C., Burns, K.H., Neretti, N., and Rodić, N. (2016) Genome-wide characterization of humanL1 antisense promoter-driven transcripts. BMC Genomics, 17:463).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Camilla Andersen, Claire Katrine Beier Holden, Asli Silahtaroglu, Elisabeth Larsen, Mads Bak (Departments of Odontology and Cellular and Molecular Medicine, Copenhagen Center for Glycomics, University of Copenhagen) and Katja Dahlgaard, (Department of Science and Environment, Roskilde University, Denmark) for excellent technical assistance. We thank Kim Madsen at CLC Bioscience for guidance and access to CLC Workbench.

Author contributions: R.P. planned and performed all ZFN targeting, IHC, FACS and live imaging experiments. L.H., M.C and C.M. planned and performed RNA-seq and ChIPSeq experiments. S.L. and J.D. performed all *in vitro* promoter analysis. J.B.H. performed Dox induction and FACS experiments. J.T.T., R.A. and L.D. contributed with writing parts of the manuscript. E.P.B. designed and planned all experiments and wrote the manuscript.

FUNDING

University of Copenhagen Excellence Programme for Interdisciplinary Research [CDO2016]; The Danish Council for Independent Research [4004-00140B], Novo Nordisk Foundation, Danish National Research Foundation [DNRF107], FEDER—Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020—Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020; Portuguese funds through FCT—Fundação para a Ciência e a Tecnologia/Ministério da Ciência, Tecnologia e Inovação [POCI-01-0145-FEDER-007274, POCI-01-0145-FEDER-016390]. Funding for open access charge: Danish National Research Foundation [DNRF107].

Conflict of interest statement. None declared.

REFERENCES

1. McClintock,B. (1950) the origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U.S.A.*, **36**, 344–355.
2. Goth,R. and Rajewsky,M.F. (1972) Ethylation of nucleic acids by ethylnitrosourea-l-14C fetal and adult Rat1. *Cancer Res.*, **32**, 1501–1505.
3. Hinton,C. and Whittinghill,M. (1950) The distribution of X-ray induced crossovers from curly inversion heterozygotes of Drosophila melanogaster females. *Proc. Natl. Acad. Sci. U.S.A.*, **36**, 552–558.
4. Capecchi,M.R. (1980) High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. *Cell*, **22**, 479–488.
5. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
6. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.a. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
7. Geurts,A.M., Cost,G.J., Freyvert,Y., Zeitler,B., Jeffrey,C., Choi,V.M., Jenkins,S.S., Wood,A., Cui,X., Meng,X. *et al.* (2010) Knockout rats produced using designed zinc finger nucleases. *Science*, **325**, 2009–2011.
8. Fire,A., Xu,S., Montgomery,M.K., Kostas,S.A., Driver,S.E. and Mello,C.C. (1998) Potent and specific genetic interference by

double-stranded RNA in Caenorhabditis elegans. *Nature*, **391**, 806–811.
9. Liu,J., Deutsch,U., Fung,I. and Lobe,C.G. (2014) Conditional and inducible transgene expression in endothelial and hematopoietic cells using Cre/loxP and tetracycline-off systems. *Exp. Ther. Med.*, **8**, 1351–1356.
10. Kuhn,R., Schwenk,F., Aguet,M. and Rajewsky,K. (1995) Inducible gene targeting in mice. *Science*, **269**, 1427–1428.
11. Braselmann,S., Graninger,P. and Busslinger,M. (1993) A selective transcriptional induction system for mammalian cells based on Gal4-estrogen receptor fusion proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **90**, 1657–1661.
12. Kennedy,M.J., Hughes,R.M., Peteya,L.A., Schwartz,J.W., Ehlers,M.D. and Tucker,C.L. (2010) Rapid blue-light – mediated induction of protein interactions in living cells. *Nat. Methods*, **7**, 12–16.
13. No,D., Yao,T.P. and Evans,R.M. (1996) Ecdysone-inducible gene expression in mammalian cells and transgenic mice. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 3346–3351.
14. Unger,B. and Hillen,W. (1984) Nucleotide sequence of the repressor gene of the RAl tetracycline resistance determinant: structural and functional comparison with three related Tet repressor genes Bernhard. *Nucleic Acids Res.*, **12**, 7693–7703.
15. Postle,K., Nguyen,T.T. and Bertrand,K.P. (1984) Nucleic acids research. *Nucleic Acids Res.*, **12**, 4849–4863.
16. Gossen,M. and Bujard,H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 5547–5551.
17. Gossen,M., Freundlieb,S., Bender,G., Müller,G., Hillen,W. and Bujard,H. (1995) Transcriptional activation by tetracyclines in mammalian cells. *Science*, **268**, 1766–1769.
18. Das,A.T., Zhou,X., Metz,S.W., Vink,M.A. and Berkhout,B. (2016) Selecting the optimal Tet-On system for doxycycline-inducible gene expression in transiently transfected and stably transduced mammalian cells. *Biotechnol. J.*, **11**, 71–79.
19. Urlinger,S., Baron,U., Thellmann,M., Hasan,M.T., Bujard,H. and Hillen,W. (2000) Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 7963–7968.
20. Freundlieb,S., Schirra-Müller,C. and Bujard,H. (1999) A tetracycline controlled activation/repression system with increased potential for gene transfer into mammalian cells. *J. Gene Med.*, **1**, 4–12.
21. Loew,R., Heinz,N., Hampf,M., Bujard,H. and Gossen,M. (2010) Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnol.*, **10**, 81.
22. Heinz,N., Schambach,A., Galla,M., Maetzig,T., Baum,C., Loew,R. and Schiedlmeier,B. (2011) Retroviral and transposon-based tet-regulated all-in-one vectors with reduced background expression and improved dynamic range. *Hum. Gene Ther.*, **22**, 166–176.
23. Das,A.T., Liliane,T. and Berkhout,B. (2016) Tet-On systems for doxycycline-inducible gene expression. *Curr. Gene Ther.*, **16**, 156–167.
24. Moullan,N., Mouchiroud,L., Houtkooper,R.H., Auwerx,J., Moullan,N., Mouchiroud,L., Wang,X., Ryu,D., Williams,E.G. and Mottis,A. (2015) Tetracyclines disturb mitochondrial function across eukaryotic models: a call for caution in biomedical article tetracyclines disturb mitochondrial function across eukaryotic models: a call for caution in biomedical research. *Cell Rep.*, **10**, 1681–1691.
25. Bensinger,S.J., Graeber,T.G. and Christofk,H.R. (2013) Doxycycline alters metabolism and proliferation of human cell lines. *PLoS One*, **8**, 8–14.
26. Brüning,A., Brem,G.J., Vogel,M. and Mylonas,I. (2013) Tetracyclines cause cell stress-dependent ATF4 activation and mTOR inhibition. *Exp. Cell Res.*, **320**, 281–289.
27. Zhou,X., Vink,M., Klaver,B., Berkhout,B. and Das,A.T. (2006) Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene Ther.*, **13**, 1382–1390.
28. Suh,E., Chen,L., Taylor,J. and Traber,P.G. (1994) A homeodomain protein related to caudal regulates intestine-specific gene transcription. *Mol. Cell. Biol.*, **14**, 7340–7351.
29. Duda,K., Lonowski,L.a., Kofoed-Nielsen,M., Ibarra,A., Delay,C.M., Kang,Q., Yang,Z., Pruett-Miller,S.M., Bennett,E.P., Wandall,H.H. *et al.* (2014) High-efficiency genome editing via 2A-coupled co-expression of fluorescent proteins and zinc finger nucleases or CRISPR/Cas9 nickase pairs. *Nucleic Acids Res.*, doi:10.1093/nar/gku251.
30. Maresca,M., Lin,V.G., Guo,N. and Yang,Y. (2013) Obligate Ligation-Gated Recombination (ObLiGaRe): custom-designed nucleases mediated targeted integration through nonhomologous end joining. *Genome Res.*, doi:10.1101/gr.145441.112.
31. Wandall,H.H., Rumjantseva,V., Sørensen,A.L.T., Patel-Hett,S., Josefsson,E.C., Bennett,E.P., Italiono,J.E., Henrik,C., H,H.J. and Hoffmeister,K.M. (2012) The origin and function of platelet glycosyltransferases Glycosyltransferase assays Immunoblot analysis Flow cytometric analysis of platelet glycosyltransferases. *Blood*, **120**, 626–635.
32. Yang,Z., Steentoft,C., Hauge,C., Hansen,L., Thomsen,a. L., Niola,F., Vester-Christensen,M.B., Frodin,M., Clausen,H., Wandall,H.H. *et al.* (2015) Fast and sensitive detection of indels induced by precise gene targeting. *Nucleic Acids Res.*, **43**, e59.
33. Lonowski,L.A., Narimatsu,Y., Riaz,A., Delay,C.E., Yang,Z., Niola,F., Duda,K., Ober,E.A., Clausen,H., Wandall,H.H. *et al.* (2017) Genome editing using FACS enrichment of nuclease-expressing cells and indel detection by amplicon analysis. *Nat. Protoc.*, **12**, 581–603.
34. Coskun,M., Troelsen,J.T. and Nielsen,O.H. (2011) The role of CDX2 in intestinal homeostasis and inflammation. *Biochim. Biophys. Acta*, **1812**, 283–289.
35. Ji,H., Jiang,H., Ma,W., Johnson,D.S., Myers,R.M. and Wong,W.H. (2008) An integrated system CisGenome for analyzing ChIP-chip and ChIP-seq data. *Nat. Biotechnol.*, **26**, 1293–1300.
36. Gao,N., White,P. and Kaestner,K.H. (2010) Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. *Dev. Cell*, **16**, 588–599.
37. Mesquita,P., Jonckheere,N., Almeida,R., Ducourouble,M.P., Serpa,J., Silva,E., Pigny,P., Santos Silva,F., Reis,C., Silberg,D. *et al.* (2003) Human MUC2 mucin gene is transcriptionally regulated by Cdx homeodomain proteins in gastrointestinal carcinoma cell lines. *J. Biol. Chem.*, **278**, 51549–51556.
38. Boyd,M., Hansen,M., Jensen,T.G.K., Perearnau,A., Olsen,A.K., Bram,L.L., Bak,M., Tommerup,N., Olsen,J. and Troelsen,J.T. (2010) Genome-wide analysis of CDX2 binding in intestinal epithelial cells (Caco-2). *J. Biol. Chem.*, **285**, 25115–25125.
39. Bennett,E.P., Mandel,U., Clausen,H., Gerken,T.a, Fritz,T.a and Tabak,L.a (2012) Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*, **22**, 736–756.
40. Salari,K., Spulak,M.E., Cuff,J., Forster,A.D., Giacomini,C.P. and Huang,S. (2012) CDX2 is an amplified lineage-survival oncogene in colorectal cancer Keyan. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 3196–3205.
41. Rutzky,L.P., Kaye,C.I., Siciliano,M.J., Chao,M. and Kahan,B.D. (1980) Longitudinal karyotype and genetic signature analysis of cultured human colon adenocarcinoma cell lines LS180 and LS174T. *Cancer Res.*, **40**, 1443–1448.
42. Göttke,M.S., Keller,K., Belley,A., Garcia,R.-M., Hollingsworth,M.A., Mack,D.R. and Chadee,K. (1998) Functional heterogeneity of colonic adenocarcinoma mucins for inhibition of Entamoeba histolytica adherence to target cells. *J. Eukaryot. Microbiol.*, **45**, 17S–23S.
43. Bu,X.D., Li,N., Tian,X.Q. and Huang,P.L. (2011) Caco-2 and LS174T cell lines provide different models for studying mucin expression in colon cancer. *Tissue Cell*, **43**, 201–206.
44. Heath,J.K., White,S.J., Johnstone,C.N., Catimel,B., Simpson,R.J., Moritz,R.L., Tu,G.F., Ji,H., Whitehead,R.H., Groenen,L.C. *et al.* (1997) The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 469–474.
45. Pinto,R., Barros,R., Pereira-Castro,I., Mesquita,P., da Costa,L.T., Bennett,E.P., Almeida,R. and David,L. (2015) CDX2 homeoprotein is involved in the regulation of ST6GalNAc-I gene in intestinal metaplasia. *Lab. Invest.*, **95**, 1–10.
46. Sakuma,K., Aoki,M. and Kannagi,R. (2012) Transcription factors c-Myc and CDX2 mediate E-selectin ligand expression in colon cancer cells undergoing EGF / bFGF-induced epithelial—mesenchymal transition. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 7776–7781.

47. Isshiki,S., Kudo,T., Nishihara,S., Ikehara,Y., Togayachi,A., Furuya,A., Shitara,K., Kubota,T., Watanabe,M., Kitajima,M. *et al.* (2003) Lewis type 1 antigen synthase (β3Gal-T5) is transcriptionally regulated by homeoproteins. *J. Biol. Chem.*, **278**, 36611–36620.
48. Mitchelmore,C., Troelsen,J.T., Spodsberg,N., Sjöström,H. and Norén,O. (2000) Interaction between the homeodomain proteins Cdx2 and HNF1alpha mediates expression of the lactase-phlorizin hydrolase gene. *Biochem. J.*, **346**, 529–535.
49. Hansen,L., Lind-thomsen,A., Joshi,H.J., Pedersen,N.B., Have,C.T., Kong,Y., Wang,S., Sparso,T., Grarup,N., Vester-christensen,M.B. *et al.* (2015) A glycogene mutation map for discovery of diseases of glycosylation. *Glycobiology*, **25**, 211–224.
50. Rizzino,A. (2009) Sox2 and Oct-3/4: a versatile pair of master regulators that orchestrate the self-renewal and pluripotency of embryonic stem cells. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, **1**, 228–236.
51. Yang,Z., Wang,S., Halim,A., Schulz,M.A., Frodin,M., Rahman,S.H., Vester-Christensen,M.B., Behrens,C., Kristensen,C., Vakhrushev,S.Y. *et al.* (2015) Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat. Biotechnol.*, doi:10.1038/nbt.3280.
52. Bahr,S., Cortner,L., Ladley,S. and Borgschulte,T. (2013) Evaluating the effect of chromosomal context on zinc finger nuclease efficiency. *BMC Proc.*, **7**, P3.
53. Mansour,W.Y., Schumacher,S., Rosskopf,R., Rhein,T., Schmidt-Petersen,F., Gatzemeier,F., Haag,F., Borgmann,K., Willers,H. and Dahm-Daphi,J. (2008) Hierarchy of nonhomologous end-joining, single-strand annealing and gene conversion at site-directed DNA double-strand breaks. *Nucleic Acids Res.*, **36**, 4088–4098.
54. Cristea,S., Freyvert,Y., Santiago,Y., Holmes,M.C., Urnov,F.D., Gregory,P.D. and Cost,G.J. (2013) In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol. Bioeng.*, **110**, 871–880.
55. Yang,Z., Wang,S., Halim,A., Schulz,M.A., Frodin,M., Rahman,S.H., Vester-christensen,M.B., Behrens,C., Kristensen,C., Vakhrushev,S.Y. *et al.* (2015) Engineered CHO cells for production of diverse , homogeneous glycoproteins. *Nat. Biotechnol.*, **33**, 842–844.
56. Inui,M., Miyado,M., Igarashi,M., Tamano,M., Kubo,A., Yamashita,S., Asahara,H., Fukami,M. and Takada,S. (2014) Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Sci. Rep.*, **4**, 5396.
57. Lee,J.S., Grav,L.M., Pedersen,L.E., Lee,G.M. and Kildegaard,H.F. (2016) Accelerated homology-directed targeted integration of transgenes in Chinese hamster ovary cells via CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. Bioeng.*, **9999**, 1–17.
58. He,X., Tan,C., Wang,F., Wang,Y., Zhou,R., Cui,D., You,W., Zhao,H., Ren,J. and Feng,B. (2016) Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res.*, **44**, e85.
59. Dow,L.E., Fisher,J., O'Rourke,K.P., Muley,A., Kastenhuber,E.R., Livshits,G., Tschaharganeh,D.F., Socci,N.D. and Lowe,S.W. (2015) Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.*, **33**, 390–394.
60. Wang,J., Exline,C.M., DeClercq,J.J., Llewellyn,G.N., Hayward,S.B., Li,P.W.-L., Shivak,D.A., Surosky,R.T., Gregory,P.D., Holmes,M.C. *et al.* (2015) Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nat. Biotechnol.*, **33**, 1256–1263.
61. Cai,Y., Laustsen,A., Zhou,Y., Sun,C., Anderson,M.V., Li,S., Uldbjerg,N., Luo,Y., Jakobsen,M.R. and Mikkelsen,J.G. (2016) Targeted, homology-driven gene insertion in stem cells by ZFN-loaded ' all-in-one ' lentiviral vectors. *Elife*, e12213.
62. Coluccio,A., Miselli,F., Lombardo,A., Marconi,A., Malagoli Tagliazucchi,G., Gonçalves,M.a, Pincelli,C., Maruggi,G., Del Rio,M., Naldini,L. *et al.* (2013) Targeted gene addition in human epithelial stem cells by zinc-finger nuclease-mediated homologous recombination. *Mol. Ther.*, **21**, 1695–1704.
63. Moehle,E.a, Rock,J.M., Lee,Y.-L., Jouvenot,Y., DeKelver,R.C., Gregory,P.D., Urnov,F.D. and Holmes,M.C. (2007) Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 3055–3060.
64. Yamamoto,Y., Bliss,J. and Gerbi,S.A. (2015) Whole organism genome editing: targeted large DNA insertion via ObLiGaRe nonhomologous end-joining in vivo capture. *G3 (Bethesda).*, **5**, 1843–1847.
65. Yang,Z., Wang,S., Halim,A., Schulz,M.A., Frodin,M., Rahman,S.H., Vester-Christensen,M.B., Behrens,C., Kristensen,C., Vakhrushev,S.Y. *et al.* (2015) Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat. Biotechnol.*, **33**, 2014–2017.
66. Brown,M.S., Goldstein,J.L., Arnold,W. and Perdue,S. (1974) Expression of the familial hypercholesterolemia gene in heterozygotes: mechanism for a dominant disorder in man. *Science*, **185**, 61–63.
67. Van Cong,N., Aubert,J.P., Gross,M.S., Porchet,N., Degand,P. and Frézal,J. (1990) Assignment of human tracheobronchial mucin gene(s) to 11p15 and a tracheobronchial mucin-related sequence to chromosome 13. *Hum. Genet.*, **86**, 167–172.
68. Gerard,C., Eddy,R.L. and Shows,T.B. (1990) The core polypeptide of cystic fibrosis tracheal mucin contains a tandem repeat structure: evidence for a common mucin in airway and gastrointestinal tissue. *J. Clin. Invest.*, **86**, 1921–1927.
69. Velcich,A., Palumbo,L., Selleri,L., Evans,G. and Augenlicht,L. (1997) Organization and regulatory aspects of the human intestinal mucin gene (MUC2) locus. *J. Biol. Chem.*, **272**, 7968–7976.
70. Wheeler,D.a., Gibbs,R.a., Muzny,D.M., Bainbridge,M.N., Chang,K., Dinh,H.H., Drummond,J.a., Fowler,G., Kovar,C.L., Lewis,L.R. *et al.* (2012) Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, **487**, 330–337.
71. Sakamoto,J., Kojima,H., Kato,J., Hamashima,H. and Suzuki,H. (2000) Organ-specific expression of the intestinal epithelium-related antigen A33, a cell surface target for antibody-based imaging and treatment in gastrointestinal cancer. *Cancer Chemother. Pharmacol.*, **46**(Suppl), S27–S32.
72. Ciprotti,M., Chong,G., Gan,H.K., Chan,A., Murone,C., MacGregor,D., Lee,F.-T., Johns,T.G., Heath,J.K., Ernst,M. *et al.* (2014) Quantitative intratumoural microdistribution and kinetics of (131)I-huA33 antibody in patients with colorectal carcinoma. *EJNMMI Res.*, **4**, 1–10.
73. Steentoft,C., Vakhrushev,S.Y., Joshi,H.J., Kong,Y., Vester-Christensen,M.B., Schjoldager,K.T.-B.G., Lavrsen,K., Dabelsteen,S., Pedersen,N.B., Marcos-Silva,L. *et al.* (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO J.*, **32**, 1478–1488.
74. Kong,Y., Joshi,H.J., Schjoldager,K.T.-B.G., Madsen,T.D., Gerken,T.a, Vester-Christensen,M.B., Wandall,H.H., Bennett,E.P., Levery,S.B., Vakhrushev,S.Y. *et al.* (2014) Probing polypeptide GalNAc-transferase isoform substrate specificities by in vitro analysis. *Glycobiology*, **0**, 1–11.
75. Horton,J.D., Goldstein,J.L. and Brown,M.S. (2002) SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J. Clin. Invest.*, **109**, 1125–1131.
76. Shimano,H. (2001) Sterol regulatory element-binding proteins (SREBPs): transcriptional regulators of lipid synthetic genes. *Prog. Lipid Res.*, **40**, 439–452.